DOCUMENT RESUME

ED 430 997 TM 029 823

AUTHOR Shen, Linjun
TITLE A Multilevel Assessment of Differential Item Functioning.
PUB DATE 1999-04-00
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Evaluation Methods; Higher Education; *Item Bias; Medical Students; Osteopathy; *Test Items
IDENTIFIERS *Hierarchical Linear Modeling; Logistic Regression; *Multilevel Analysis

ABSTRACT

A multilevel approach was proposed for the assessment of differential item functioning and compared with the traditional logistic regression approach. Data from the Comprehensive Osteopathic Medical Licensing Examination for 2,300 freshman osteopathic medical students were analyzed. The multilevel approach used three-level hierarchical generalized linear models. The software HLM for Windows executed the hierarchical linear model (HLM) analysis and the Statistical Analysis System Proc Logistic was used for the conventional logistic regression analysis. It was not surprising to see that HLM was more conservative in identifying DIF in this study. The study demonstrates that the multilevel approach to DIF is meaningful and the results of its use appear reasonable. Implications for use of the two approaches are discussed. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 997

A Multilevel Assessment of Differential Item Functioning

Linjun Shen, Ph.D., MPH

National Board of Osteopathic Medical Examiners

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Linjun Shen

A paper presented at the 1999 AERA annual conference

Montreal, Canada

TM029823

BEST COPY AVAILABLE

A Multilevel Assessment of Differential Item Functioning

Linjun Shen, Ph.D., MPH

National Board of Osteopathic Medical Examiners

Traditional methods of differential item functioning (DIF) analysis, such as stratified Mantel-Haenszel methods[1], logistic regression analysis[2], or conventional Rasch model approach[3], all assume the differentiating factors function in the same pattern for examinees with the same characteristics. Very often, examinees are nested in different organizations such as classes, schools, or countries. The behavior of differentiating factors may vary among such organizations. For example, in some schools, boys may be good at certain topics of math, in other schools, which are conscious about the gender effects on learning math, this may not be true. Without considering the impact of nesting variables, estimation of DIF may not be sufficient and understanding of the effects of DIF may not be adequate.

Assessment of DIF needs a multilevel perspective. The purpose of this study was to propose a multilevel approach to DIF and compare this approach with traditional logistic regression approach.

Methods

Instruments and subjects

This study used the data from the Comprehensive Osteopathic Medical Licensing Examination (COMLEX) June 1998 Level 1 examination developed by the National Board of Osteopathic Medical Examiners (NBOME). A total of 2300 freshman osteopathic medical students from all the 17 osteopathic schools took the exam. The number of students of each school ranged from 56 to 240.

The exam had 800 multiple choice items with a KR-20 reliability of .92. To explore the multilevel approach and demonstrate the differences of this approach from the conventional logistic regression, this study randomly selected 15 items for analysis from the discipline Osteopathic Principles and Practices (OPP), curricula of which were believed varying among osteopathic medical schools.

Modeling

This multilevel approach used a three-level hierarchical generalized linear models (HGLM)[4]. The level-1 model was the measurement model:

$$\eta_{ijk} = \beta_{0jk} + \beta_{1jk}X_{1jk} + \beta_{2jk}X_{2jk} + \cdots \beta_{qjk}X_{qjk} + \cdots + \beta_{14jk}X_{14jk}$$

This model was a logistic regression model. η_{ijk} was a log odds of person j of school k's success in item i. X_{1ijk} through X_{14ijk} had a value of 0 or 1 identifying item i. β_{qjk} was the ability of person j in school k on item q. β_{0jk} represented the ability level on the 15th item when all of the X_{qijk} had a value of 0.

The Level-2 model estimated the person ability on item q at the Level-1 by adjusting individual examinees' overall achievement level on the subject OPP since the chance to succeed on any of the OPP item was certainly the function of that person's overall knowledge level on that discipline. The overall achievement level of OPP was generated by the Rasch model based on all 90 OPP items in the exam. In the Level-2 model, the person measure of OPP ability was centered around school mean into $\overline{OPP_k}$ so that it represented the mean achievement level on OPP for examinees in school k.

$$\beta_{qjk}=\gamma_{q0k}+\gamma_{q1k}\overline{(OPP)}_k+\upsilon_{jk}$$

Since $\overline{OPP_k}$ was the school mean, conditioned on $\overline{OPP_k}$, γ_{q0k} became the unique ability on item q for average students in school k.

The Level-3 models captured the school effects on the Level-2 random coefficients:

$$\gamma_{q0k}=\pi_{q00}+\nu_{q0k}$$

$$\gamma_{q1k}=\pi_{q01}+\nu_{q1k}$$

In the Level-3 model, π_{q00} was the grand mean of all 17 school and ν_{qok} was the deviation from the grand mean for school k. The parameter of interest was ν_{qok}. If Chi-square test shows it significantly different from zero, the implication would be that γ_{q0k}, the unique person ability on item q of average students in school k was significantly different from that of other schools. This would signify the DIF of item q by schools.

The above multilevel approach was compared with the following single level logistic regression approach. The notations of the following model were independent from the models above.

$$\eta_{ij}=\beta_0+\beta_1 W_1+\beta_2 W_2+\ldots\beta_q W_q+\ldots+\beta_{16} W_{16}+\beta_{17} OPP_j+\mu_{ij}$$

where W_q was a school indicator with a value of 0 or 1. After conditioned by the performances on OPP, any significant β_q would indicate students in school q performed differently from the baseline on the item i. This approach analyzed DIF one item a time.

There were two conceptually significant differences between the hierarchical models and the

regular logistic model. Firstly, the HLM models partitioned the overall variation among students' probability of success into the variation at school level, in this study, the differentiating factor, and variations among the probabilities on different items within the same students, and variations among students within the same schools. Therefore, the size of the actual variance across schools, in most cases, would be smaller with HLM models than with the conventional logistic model. Secondly, the impact of the confounding variable OPP in logistic regression was averaged across schools without considering its variation among schools, while in HLM models, the coefficient of OPP was modeled to have random variation at school level. Consequently, the adjustment of confounding factor in the HLM models were more precise. This can be demonstrated by a combined model of all three HLM models.

Two modeling approaches generated parameters with different meanings. Direct comparison of the parameter estimates was difficult. As a preliminary study, this paper first compared the conclusions from the two approaches in terms of the results of significant test of DIF. More specific analyses were given to the items which two approaches did not agree on. The expectations, based on the conceptualization of the differences of the two approaches, were that conventional logistic regression would overestimate the significance of the DIF and HLM approach would identify less items with significant DIF.

Results

The software HLM for Windows version 4.04 executed the HLM analysis and SAS Proc Logistic conducted the conventional logistic regression analysis. The HLM analyses found that no item had a significant within school variation. Except items 2 and item 4, none of the rest of items had a significant random effect of the slope of OPP at the Level 2. Theoretically, for the items with non-significant within school effects but significant school variations, the models could be reformulated into two-level models. For the items with non-significant within and between school variations, the models were conceptually a single level conventional logistic regression model. However, practically, the three-level formulation still resulted in some differences which could be meaningful for marginal items.

Table 1 compares the results of the two approaches. Except three items, the logistic regression found school was a significant differentiating factor for all the other items even most of them were marginally significant with significance level in the range of .055 and .07. For all those marginally significant items, only one school had a significant odds ratio. In contrast, the HLM approach signified only 5 items having DIF due to the school factor. Using the estimated parameters in the HLM residual files, 95% confidence intervals of odds ratio for items with marginal significance were calculated and listed in Table 1. Clearly, due to different conceptualization, parameters for individual schools were estimated differently by HLM and logistic regression.

The percentages of school level variance in the overall variance of η_{ijk} listed in Table 1 demonstrate that, in general, items with nonsignificant DIF due to school factor had small percentage of school level variance and vice versa. Items 7 and 9 appear to be exceptions.

Discussion

It was not surprising to see that HLM was more conservative in identifying DIF in this particular study. Variation partition and multilevel formulation of the confounding variable OPP contributed to the differences of the results of the two approaches. As results, in many cases, the actual size of DIF estimated by HLM was smaller than the conventional logistic regression.

The differences between the two approaches have a broader implication. In MH approach, the stratification variable is assumed functioning consistently across the next level variable just as in logistic regression, therefore, the arguments HLM made also apply to MH approach. However, HLM approach will make differences only when a multi-level data structure is present

Due to the limited scope, this study did not demonstrate how to model a differentiating factor at the individual level such as gender or race. However, the method presented here will still be applicable to individual level differentiating factors. The difference is that variables of gender or race need to be placed in the level 2 model instead of the level 3 model, and let the school level variables in the level 3 model estimate the random effects of the differentiating factor at the level 3.

More work need to be done to further explore and explain the differences between the two methods. Such comparison should also include MH approach. Important findings of this study are that the multilevel approach to DIF is meaningful and the results appear reasonable.

References

1. Holland, P. W., Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H. Braun (Ed.), Test Validity, Hillsdale, NJ: Lawrence Erlbaum Associates.

2. Wright, B.D., Mead, R.J., & Draba, R. (1976). Detecting and correcting test item bias with a logistic response model (Research Memorandum No. 22). Chicago, IL: University of Chicago, MESA Psychometric Laboratory.

3.Linacre, J.M., & Wright, B.D. (1986). Item bias: Mantel-Haenszel and the Rasch Model. (Research Memorandum No. 39). Chicago, IL: University of Chicago, MESA Psychometric Laboratory.

4. Bryk, A., Raudenbush, S., & Congdon, R. (1996). HLM™: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs. Scientific Software International, Inc.: Chicago, IL.

Table 1

Comparison between HLM and Conventional Logistic Regression

Item	Significance Logistic	Significance HLM	% of school level variation	95% C.I of Odd Ratio** Logistic	95% C.I of Odd Ratio** HLM
1	+	+	64	***	***
2	+*	+	80	1.03, 1.17	1.37, 2.86
3	+*	-	7	.45, .92	.36, 1.12
4	+*	+	75	1.33, 2.02	1.24, 3.56
5	+*	-	23	.28, .96	.39, 1.24
6	+	-	12	1.06,3.48	.64, 1.36
7	-	-	75	***	***
8	+*	-	58	1.02, 1.33	.72, 1.57
9	+*	-	70	1.10, 1.73	.86, 3.50
10	-	+	79	***	***
11	+*	+	34	1.14, 1.79	1.86, 3.23
12	+*	-	48	1.04, 1.36	.88, 2.57
13	+*	-	21	.66, .97	.48, 1.79
14	+*	-	12	.56, .93	.47, 1.85
15	-	-	33	***	***

* Marginally significant at .05 level. For the items with marginal significance, only one school's odds ratio was barely significant and the rest of the schools did not have a significant odds ratio.

** For marginally significant items.

*** More than one schools or none of the schools had significant odds ratio.